


Q1 2025

***FINDINGS ON
CONTEXT SENSITIVITY
EFFICIENCIES***



What's new
in this study?

This is the same data as the previous Q2 study available here but with better and deeper findings.

The main improvement centers around tokenization strategy used to tally edit distances for character based languages.

We also added 3 more languages to the study to cover a total of 20 surveyed languages.

“Machine Translation is the bane of my existence.”

Loved by researchers and enterprise buyers, hated by translators, Machine Translation has been a divisive force for the past forty years in our space.

This divisiveness is now over.

AI-powered context sensitivity in Bureau Works changes MTPE into a pleasant and dynamic experience with productivity yields that are unheard of.



MTPE

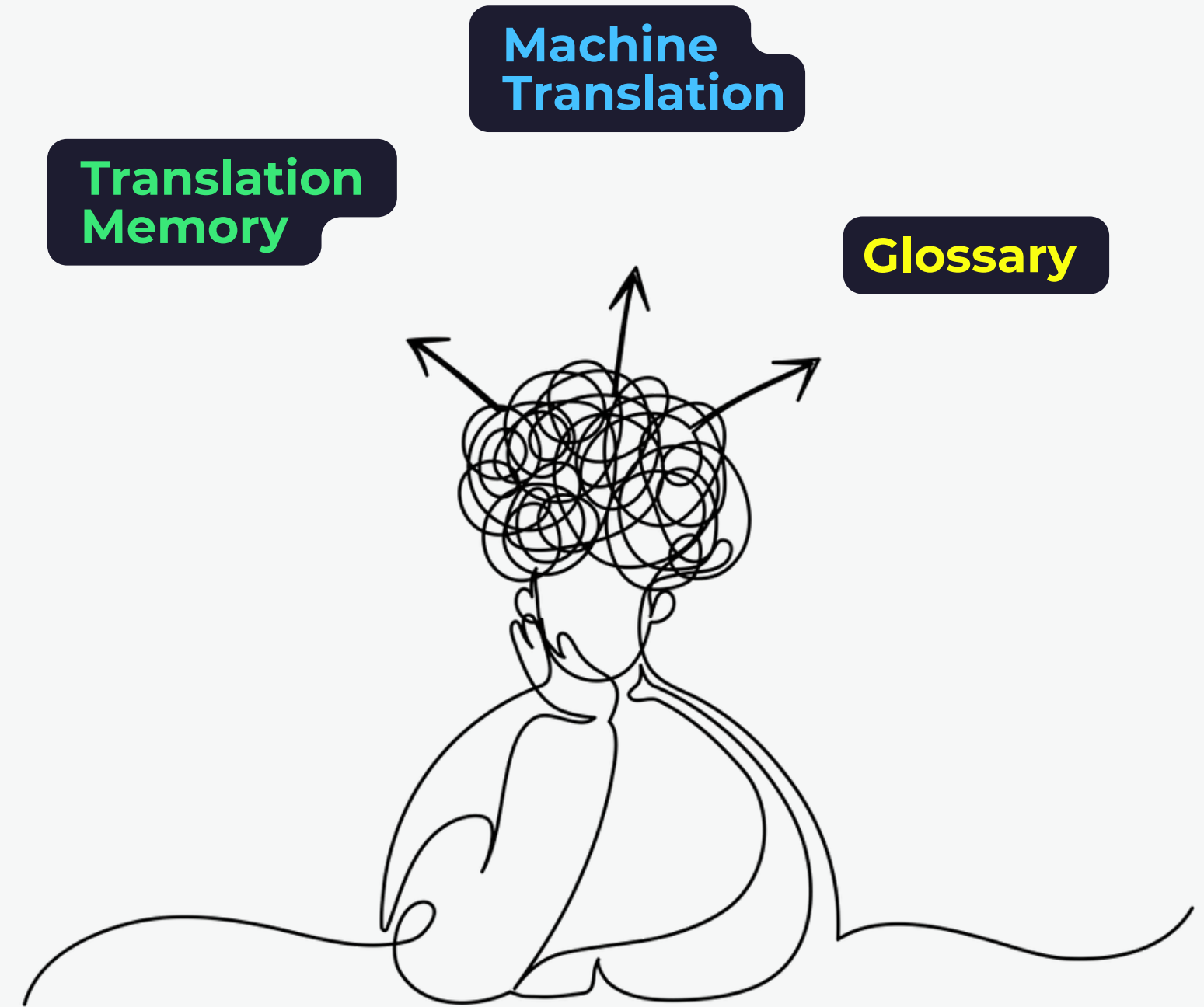
*Is Dead...
Why?*

Despite the leaps in productivity, MTPE is like life described by Thomas Hobbes: “SOLITARY, POOR, NASTY, BRUTISH, AND SHORT”

A Translator receives a file that is entirely pre-translated, and then needs to read through it trying to weed out errors, mistranslations, and inconsistencies (between MT, TM, and TB) all on their own.

Unaided, the translator’s work goes largely unleveraged and they often have to make the same changes again and again.

They work statically with a cold document that does not change or improve as a result their work.



Robbed of their sense of authorship, translators go through the mind-numbing process of reading countless lines of pre-translated text.

Their work is reduced to that of “error-sniffers”, and their pay is as well.

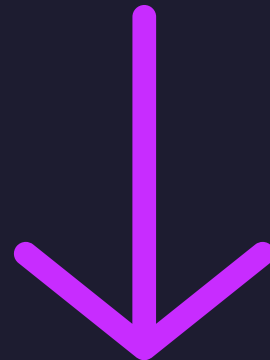
Any learning derived from their changes is only partially leveraged in training the MT model for the next batch of documents.

But this training isn't nuanced and doesn't emulate semantic awareness. Even Neural models will correlate statistical probabilities **at best.**



Context Sensitivity

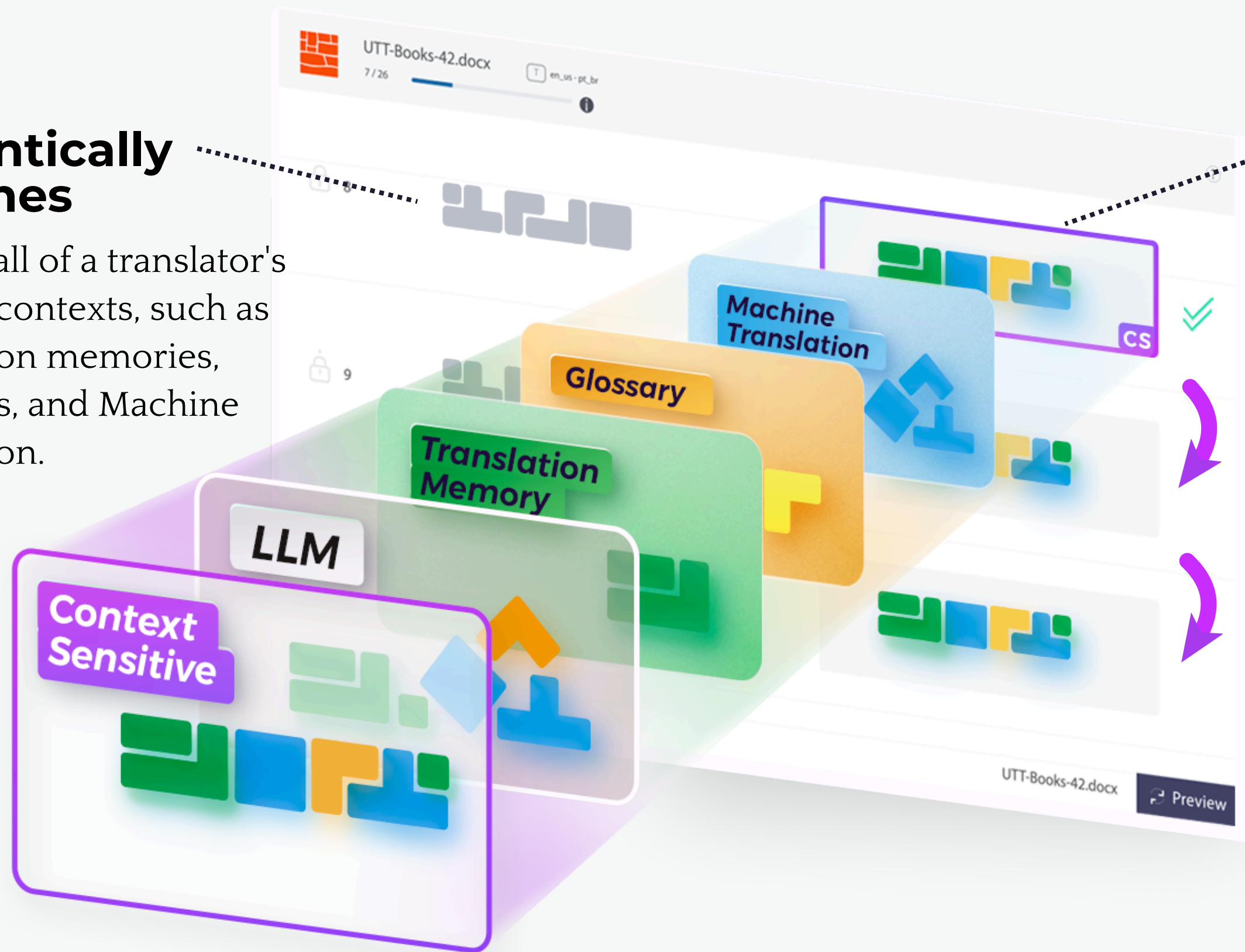
= MTPE
Reborn



Context-sensitivity uproots this legacy.

Semantically searches

through all of a translator's relevant contexts, such as Translation memories, glossaries, and Machine Translation.



Provides a Unified Translation Feed

Unlike machine translation, it offers suggestions that align with your knowledge base.

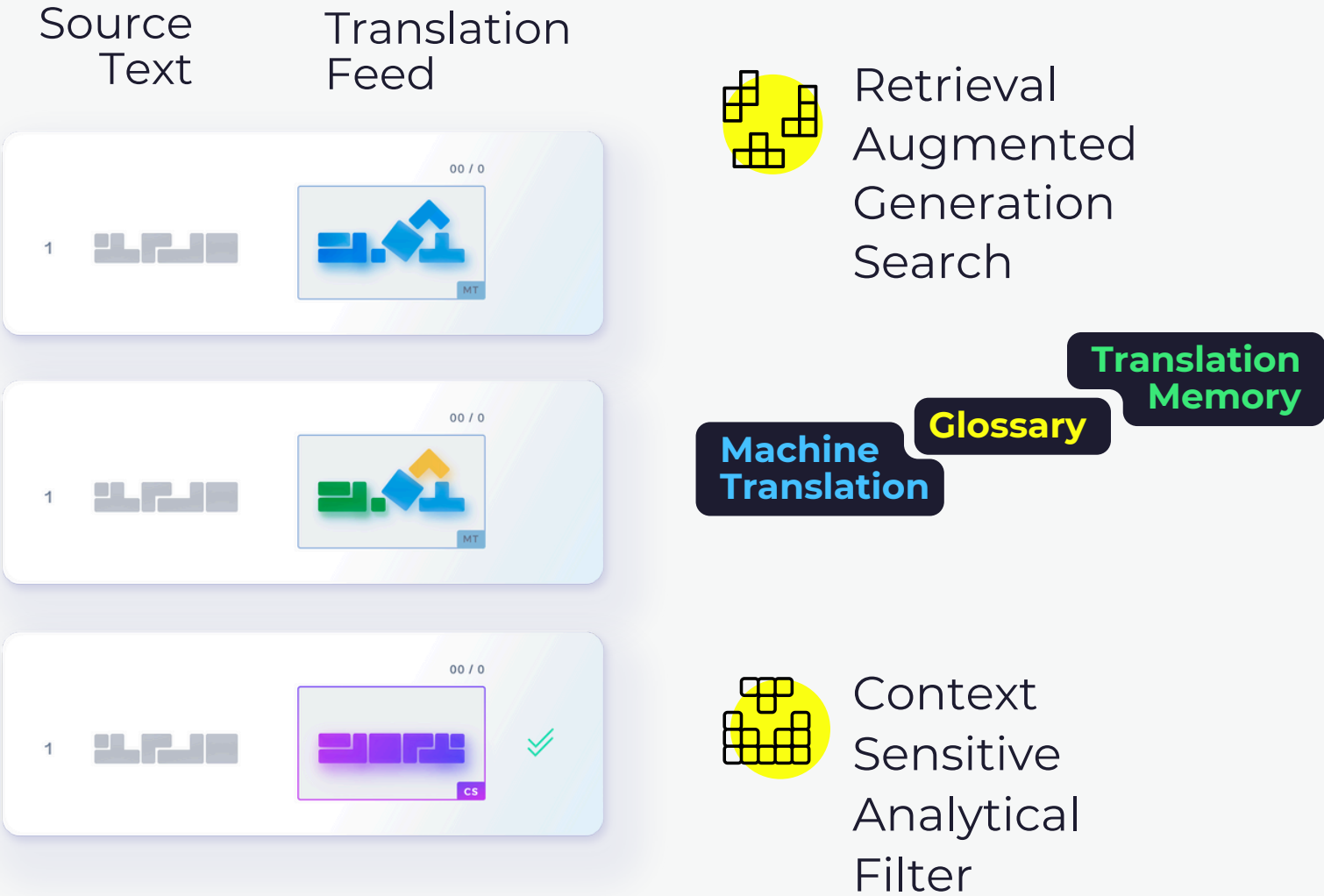
Learns instantly from translator changes

and is ready to suggest a newly adapted feed at each iteration.

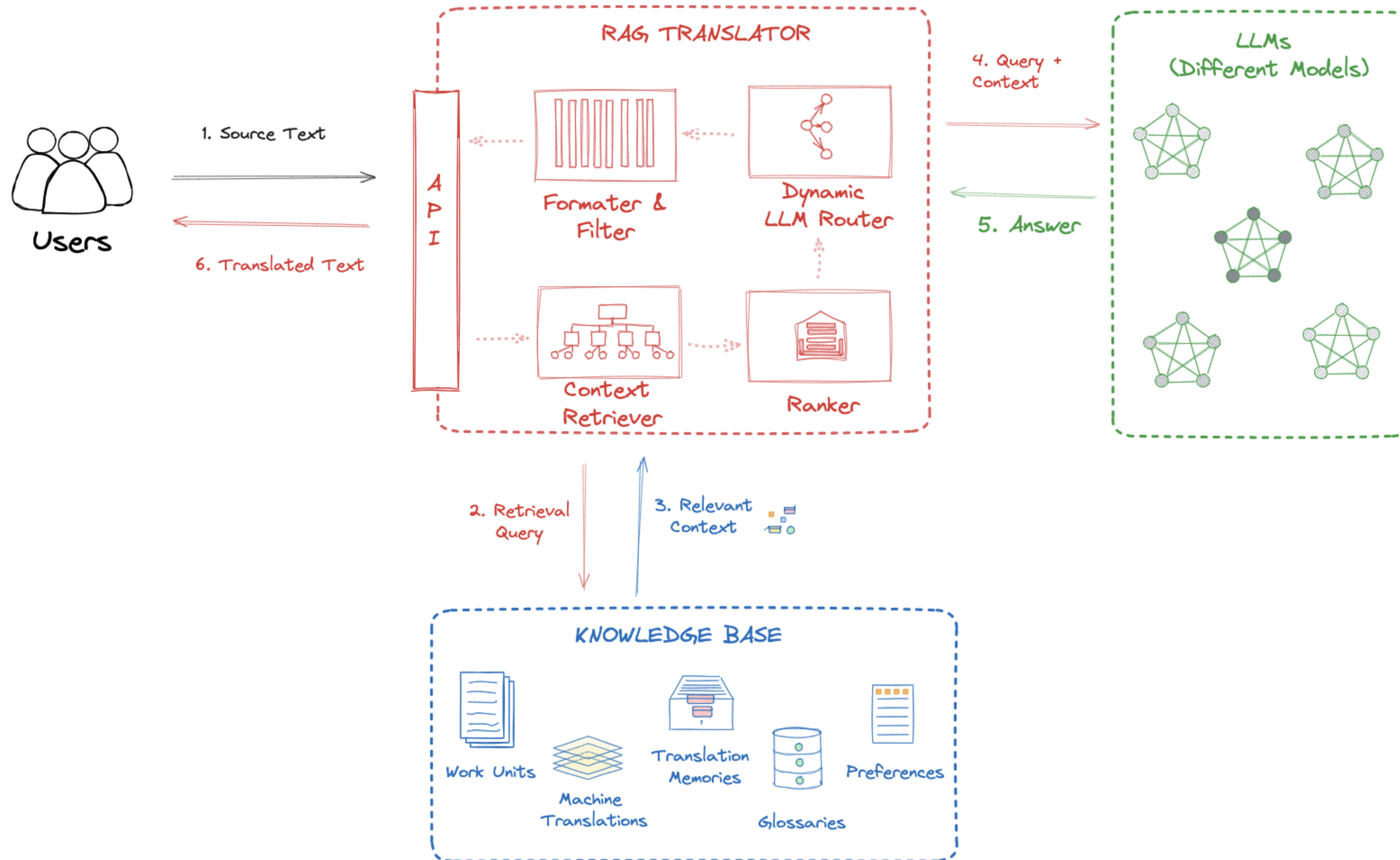
Lower-level linguistic tasks are all taken care of by the Context-Sensitive Translation:

- Ensuring **discourse agreement with TM**;
- Promoting **terminological alignment with glossary**;
- **Upgrading from fuzzy match to a perfect match**.

The linguist is left with the task of ensuring that the tone, voice and style align with their preferences. But they are not alone...



Context-sensitivity under the hood



The linguist is aided by a powerful Semantic Verifier, our "Translation Smells".

Translation Smells points translators to potential semantic issues **in runtime**:

- Additions and Omissions
- Incorrect Translations
- Gender Bias
- Subject-verb Agreement
- Awkward Phrasing
- Unnatural Sentence Structure



These are all examples of Semantic issues pointed out by Smells.

This verification not only decreases error rates, but also allows translators to move faster by pointing out potential issues regardless of human interpretation.

Some examples:



Inappropriate Register

The translation uses the informal 'te' (you) instead of the formal 'vous' which is more appropriate in this context.



Mistranslation

The phrase 'significant better quality' was translated into '更高质量' which means 'higher quality'. The correct translation would be '显著更好的质量' which means 'significantly better quality'



Terminology

The term 'environment' was translated as 'lugar' which means 'place'. A more accurate translation would be 'ambiente' which is commonly used in the context of software or systems.

The Context-Sensitive Approach is continuously learning.

The engine is taught merely by confirming segments, and terminology is self-extracted and self-aligned on the fly.

This means that your knowledge base is continuously re-nurtured without any formal training.

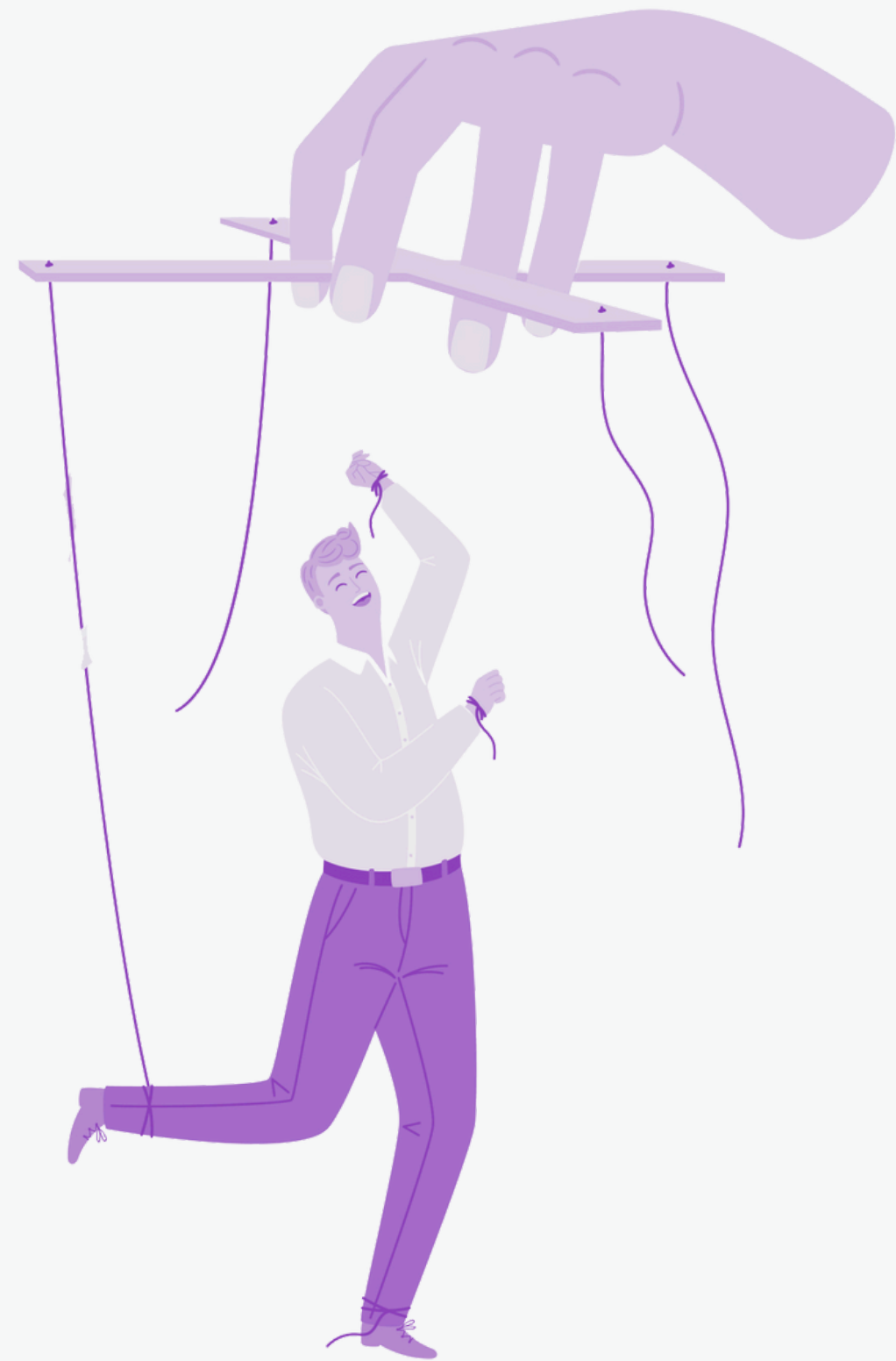
And, if at any point you would like to perform a translation memory semantic cleansing, we offer **TMill**, a tool that can cleanup and optimize anything from small to large scale TMs using Semantic Analysis.



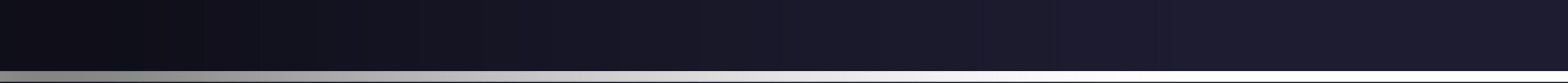
Context-sensitive technology augments translators' capabilities as authors. It allows them to translate more fluidly and expressively all while adhering to established guidelines and linguistic resources.

Translators gain the freedom to craft holistic translations without getting tripped up by consistency checks.

That's great, but what does this mean in terms of results?



our
Data



Scope

We surveyed the metadata of ~4.3 million segments between June 22 and December 23, 2023

~2.3 million were translated using traditional Neural MTPE between June 22 and Dec 23.

~2 million were translated using our Context Sensitive Approach between June 23 and Dec 23.

Microsoft Neural Machine Translation was the main engine used as a baseline in this study.

Languages

We tested a total of 20 language pairs, with English always as the source into:

Portuguese (Brazil)

Italian (Italy)

Spanish

Spanish (Spain)

Spanish (Latin America)

French (France)

French (Canada)

German (Germany)

Japanese (Japan)

Polish (Poland)

Dutch (Netherlands)

Simplified Chinese (China)

Traditional Chinese (Taiwan)

Danish (Denmark)

Russian (Russia)

Korean (Korea)

Vietnamese (Vietnam)

Thai (Thailand)

Turkish (Turkey)

Arabic (UAE)

Domains

We included content from a combination of the following domains:

Automotive	Engineering
Copywriting	Legal
Blogs	Corporate/HR
Social Media	Tech
Financial	Product
Pharmaceutical	Insurance
Banking	Restaurant Management

Approach

We examined **TER** (Translation Error Rate). TER was measured as the edit distance between the suggestion provided to the translator and what they ultimately confirmed as their entry (reference).

A 0 score means that no edits were made to the suggestion by the translator and a score of 1 means that the suggestion was entirely changed.

For further reference:

https://www.cs.umd.edu/~snoover/tercom/ter_tr.pdf

Methodology

What counts as an edit?

Any time a word was replaced, added, deleted, or moved.

The total number of edits is then divided by the total size of the reference (word count).

E.g. A 10 word suggestion with 1 word replaced will have a score of .1

Notes

Outliers with a score above 1 were treated as 1. These outliers reflect cases where the number of changes exceeds the original suggestion word count.

Translation memory leveraging in both periods was normalized. Their proportions matched in order to provide relevant numbers.

Results



from

~~18.23%~~

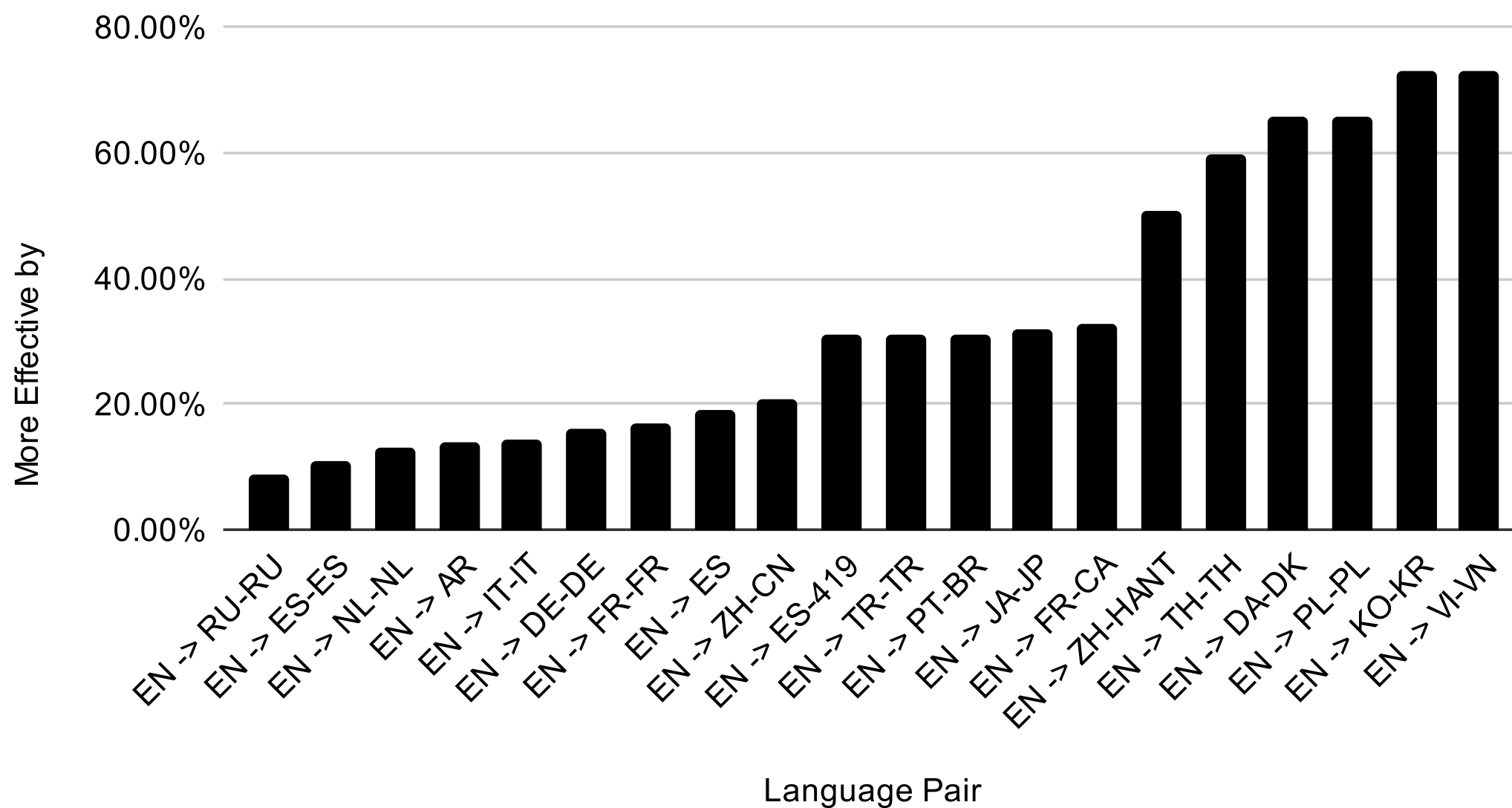
to

11.17%

Average Edit Distance

	with no MTPE (Control)	With NMT MTPE	With Context Sensitive Translate	
Language Pair	% of edits necessary from initial input to final result (TER score)	% of edits necessary from initial input to final result (TER score)	% of edits necessary from initial input to final result (TER score)	Context-Sensitive More Effective by
EN -> PT-BR	100%	0.207	0.142	31%
EN -> IT-IT	100%	0.086	0.074	14%
EN -> ES	100%	0.187	0.151	19%
EN -> ES-ES	100%	0.062	0.055	11%
EN -> ES-419	100%	0.089	0.062	31%
EN -> FR-FR	100%	0.149	0.124	17%
EN -> FR-CA	100%	0.214	0.144	33%
EN -> DE-DE	100%	0.116	0.097	16%
EN -> JA-JP	100%	.192	0.133	32%
EN -> NL-NL	100%	0.070	0.061	13%
EN -> ZH-CN	100%	.225	0.178	21%
EN -> DA-DK	100%	0.133	0.045	66%
EN -> RU-RU	100%	0.150	0.136	9%
EN -> KO-KR	100%	0.170	0.046	73%
EN -> VI-VN	100%	0.231	0.061	73%
EN -> TH-TH	100%	0.294	0.118	60%
EN -> TR-TR	100%	0.271	0.188	31%
EN -> AR	100%	.231	.199	14%
EN->PL_PL	100%	.397	.135	66%
EN_>ZH-HANT	100%	.172	.0856	51%

Context Sensitivity Improvement per Language



Context-Sensitive
was

39.28%

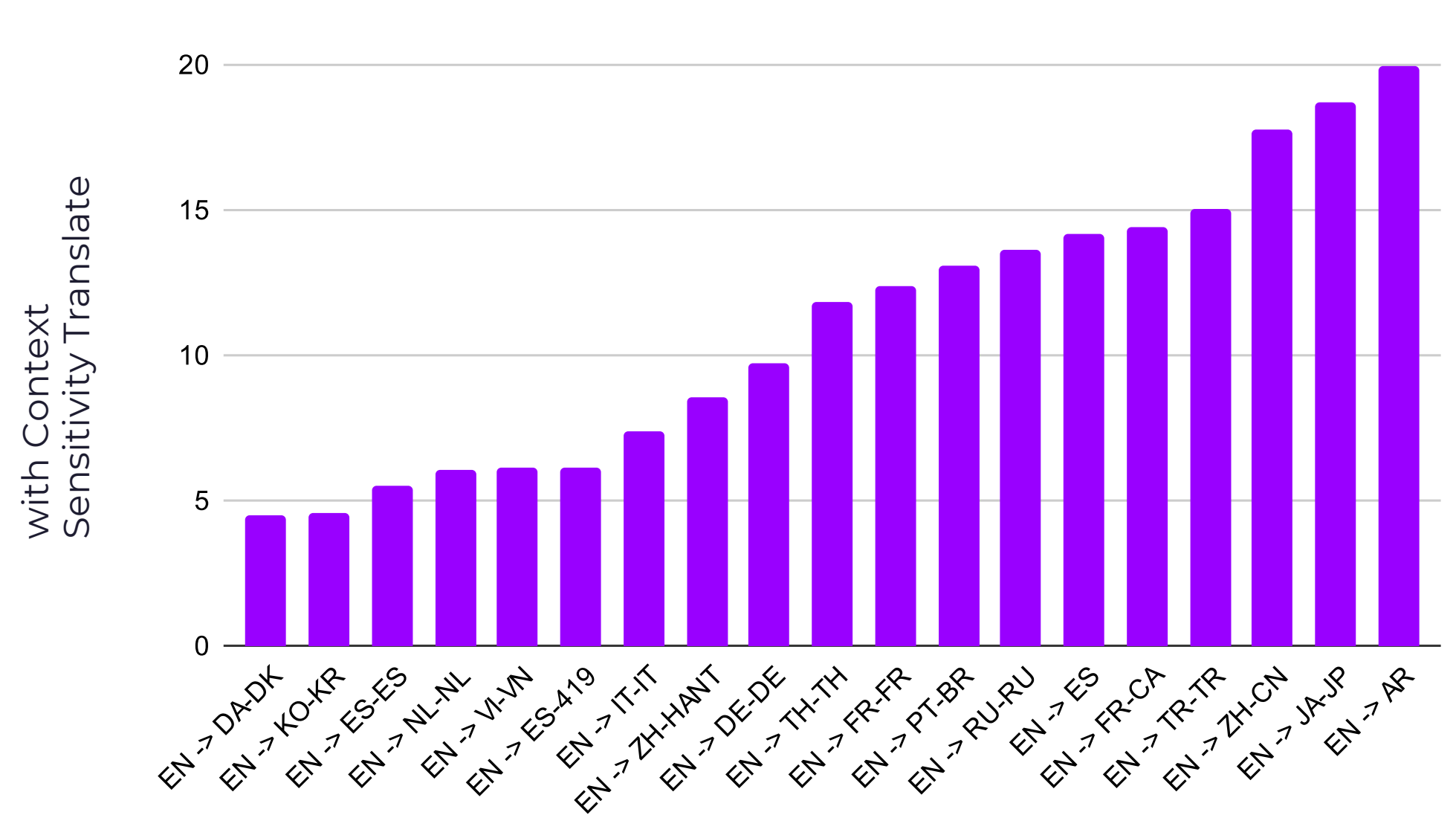
more effective
than traditional
MTPE

on average.

Context-Sensitive performance across languages using TER Score,

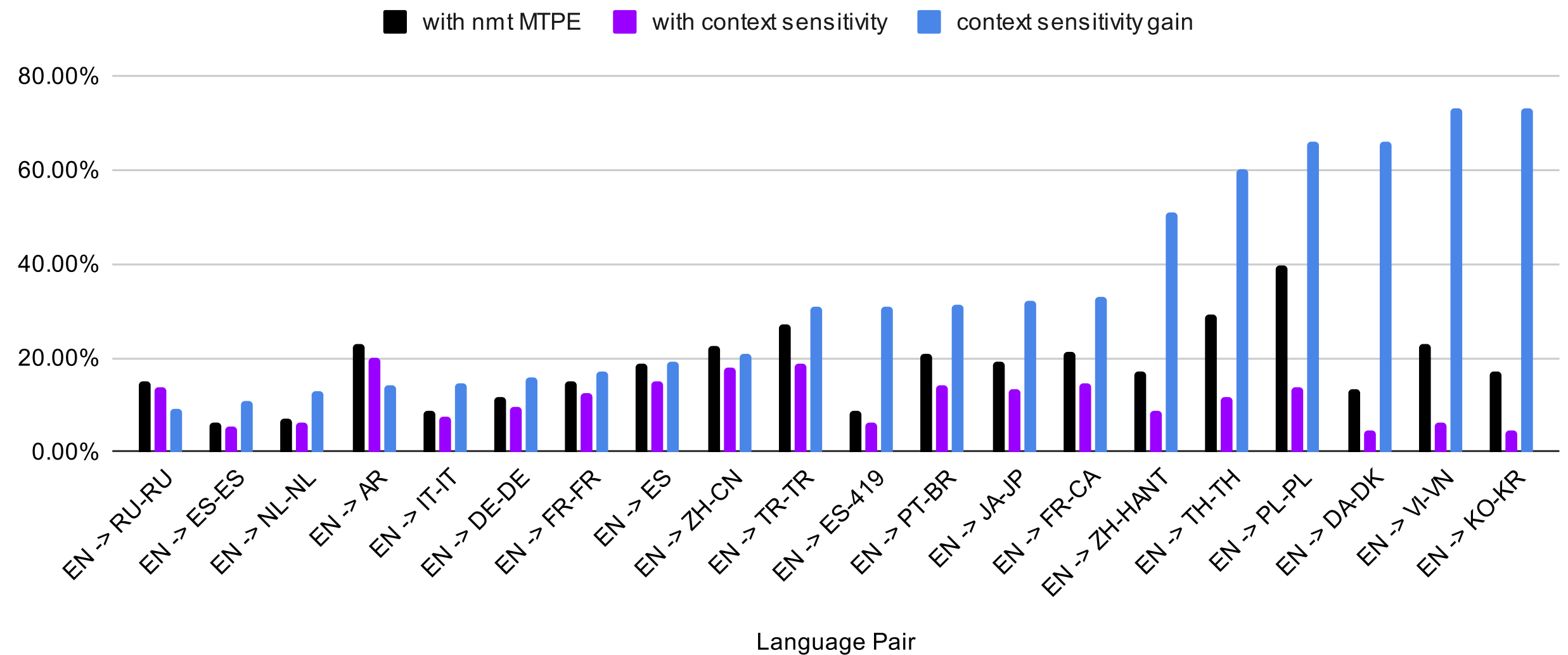
where a lower score indicates better performance.

TER Score per Language



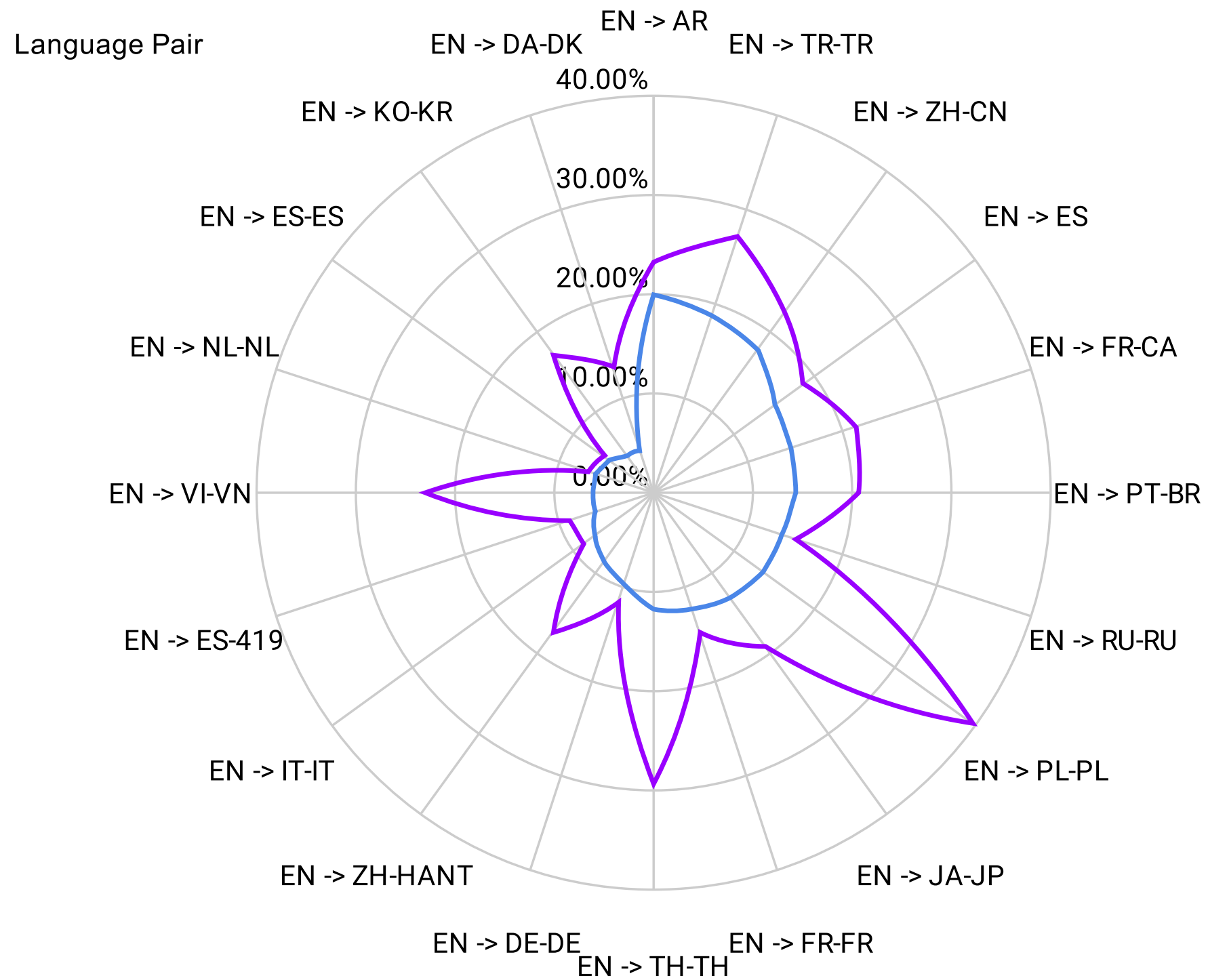
Context-Sensitive
outperforms NMT
in all languages
evaluated.

NMT MTPE versus Context Sensitivity
with Context Sensitivity Gain



NMT MTPE versus Context Sensitivity

with nmt MTPE with context sensitivity



Context Sensitive shows **more consistency across different language pairs**

While NMT performance tends to fluctuate more significantly.

Key Findings

1 - **Productivity Gains** The data clearly show that using Context Sensitivity increases translation efficiency by 22.08% on average by bringing the average TER score down from .162 to .128.

2 - **Specificity pays off** - Generic Spanish performed at 0.151 while Spanish from Spain and Latin America performed at 0.055 and 0.062 respectively, nearly 300% better.

3 - **There is significant variance** - between the top performing language (Danish) and the worst performing language (Simplified Chinese).



Past, Current and Future Hypotheses

In our previous study we raised the following questions and assumptions:

1 - Variance between languages can be associated with **tokenization methodologies** in character-based languages.

- This hypothesis proved to be true. By improving and standardizing the tokenization methodologies, non-character based languages displayed normalized edit distances for both Context Sensitive and MTPE approaches.

2 - **Translator methodologies** and training can change TER score by +/- 50%.

- This hypothesis proved true. Our latest findings on this matter show translators changing as little as 1% of the text or as much as 77%. The absolute variation of the full arrange not discounting outliers is at a whopping 7700%. This shows there is so much work to be done on training people and aligning financial incentives.

3 - TER Score is directly correlated with **Speed**.
• Not yet properly tested but initial findings to establish. this correlation

4 - Except for copywriting/transcreation, TER across specific **domains** will display variance of +/- 20.
• Not yet tested.

5 - Learning Terms will positively impact Context-Sensitive TER by at least 10%.
• Not yet tested

6 - TER Score improvement will be +/- 10% across **different NMT engines**.
• Not yet tested

7 - Translator mental well-being improves with Context Sensitivity.
• More data required

8 - Translation quality improves with Context Sensitivity.
• Not yet tested.

Conclusion

Context sensitivity continues to mature at Bureau Works and it has already produced bold results.

While we have a lot to learn, explore, and improve, our tech demonstrably delivers favorable translation outcomes.

This is only
the beginning.

